Data aggregation in wireless sensor networks: Compressing or Forecasting?

Jin Cui, Fabrice Valois

Data aggregation in Wireless Sensor Networks: Compressing or Forecasting?

Jin CUI , Fabrice VALOIS

RESEARCH REPORT
N° 8362
September 2013
Project-Teams urbanet

ISSN 0249-6399   ISRN INRIA/RR--8362--FR+ENG

Data aggregation in Wireless Sensor Networks: Compressing or Forecasting?

Jin CUI *, Fabrice VALOIS *

Project-Teams urbanet

Research Report n° 8362 — September 2013 — 26 pages

Abstract: Wireless sensor networks suffer from constrains in terms of energy, memory and computing capability. In recent years, the main challenge was to develop energy efficient solutions mainly at the MAC and network layers to increase the lifetime of the network, which spawned the development of the data aggregation. Data aggregation is the procedure of intelligently gathering information which reduce the amount of data send to the sink, this improve the network capacity. In this report, we show that data aggregation can effectively reduce the energy consuming and improve the network capacity. More, we present the state-of-the-art aggregation functions, including compressing-based and forecasting-based method; compressing aggregation focus on compress the data packets accompanied with transmitting based on spatial correlation; while forecasting aggregation tends to use mathematical model to fit the time series and predict the new value due to highly temporal correlation. We detail these two methods and characterize them respectively. We propose comparison between A-ARMA and Compressing Sensing, which are on behalf of forecasting aggregation and compressing aggregation respectively.

Key-words: data aggregation, wireless sensor networks, compressive sensing, auto-regressive moving average model

* University of Lyon, INRIA, France, INSA Lyon, CITI, F-69100, France

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Agrégation de données dans les réseaux de capteurs sans fil: la compression ou la prévision?

Résumé : Les réseaux de capteurs sans fil souffrent de contraintes en termes d'énergie, la mémoire et capacité de calcul. Au cours des derniéres annèes, le principal défi était de développer des solutions économes en énergie principalement au nieau des couches MAC et réseau pour augmenter la durée de vie du réseau, ce qui a engendré la développement de l'agrégation de données. Agrégation des données est la procédure de collecte intelligente des informaions permettant de réduire la quantité de données à envoyer de l'évier, ce améliorer la capacité du réseau. Dans ce rapport, nous montrons que l'agrégation des données peut réduire efficacement la consommation d'énergie et améliorer la capacité du réseau. De plus, nous présentons les fonctions d'agrégation state-of-the-art, y conpris méthode de compression basée sur la prévision et basée; compression accen d'agrégation sur compresse les paquets de donn'ees accompagnées de transmission sur la base de la corrélation spatiale, tandis que la prévision agrégation tend à utiliser le modèle mathématique pour s'adapter à la série temporelle et de prédire la nouvelle valeur en raison de la corrédire la nouvelle valeur en raison de la corrélation temporelle fortement. Nous détaillons ces deux méthodes et les caractériser respectivement. Nous vous proposons comparason entre A-ARMA et détection de compression, qui sont au nom de prévoir l'agrégation et la compression de l'agrégation respectivement.

Mots-clés : agrégation de données, les réseaux de capteurs sans fil, détection de compression, modèle de moyenne mobile auto-régressifs

1 Introduction

Recently wireless sensor networks (WSNs) have received more and more attention due to their potential in urban and military applications[35][22][37], including environmental monitoring, military surveillance, tracking, health care, intrusion detection, etc. Typically, WSNs have a large number of wireless sensor nodes with the ability to communicate among them and also to an external data collection point (referred as a sink). Depending of the application requirement, sensors send either periodically or on-demand for event-based the sensing data, process it and then transmit it to the sink. They are deployed to collect and report information by sending data through the WSNs to the sink. The frequency of data reporting and the number of nodes depend on the application.

Since sensor nodes are energy constrained and bandwidth constrained, it is inefficient for all the sensors to transmit the data directly to the sink. Data aggregation is defined as the process of aggregating the data from no less than one sensor to eliminate redundant transmission and provide fused information to the sink(in case of 1 sensor, we can also considerably reduce the redundant information due to the temporal correlation).

Data generated by neighbouring is often redundant and highly correlated; this is referred to as spatial correlation. Spatial aggregation usually involves the fusion of data from multiple sensors in intermediate nodes by exploiting the spatial correlation, thus the aggregated data is transmitted to the sink. Generally, sensor nodes involve in spatial correlation are in the near neighbourhood. More, consecutive data in time typically experience small variations; this is referred to as temporal correlation. Temporal correlation focus on finding the temporal correlation between the data from the same sensor in a given duration. By exploiting the temporal and/or the spatial correlation, data aggregation is a method that combines data into a high quality of information in order to reduce the data load. The less data is transmitted, the less energy and bandwidth is spent, and the more capacity of network is saved. Hence, the lifetime of WSNs is prolonged.

This report present state-of-the-art focused on aggregation function in WSNs, including two types. One is compressing-based method, i.e. compressive sensing, another is forecasting-based method. Compressing-based method focus on compress the data during the procedure of data gathering, which reduce the amount of data (or the number of packet) to achieve aggregation. Forecasting-based method tend to use mathematical model to do prediction (due to the high temporal correlation between the data) and reduce the data reporting frequency. These two methods can achieve good accuracy in sink from recovery, and many researchers do deeply research on these methods.

In section 2 we show that data aggregation can take more benefits than energy-efficient Routing and MAC protocols. Then in section 3 we discuss the compressing-based method: compressive sensing (CS), and propose the related work on CS. In section 4 we introduce the forecasting-based model and the related work. In section 5, we compare the forecasting and compressing aggregations, and give conclusion and present our future work in section 6.

2 Aggregation benefits

2.1 Some basic results

Whether data aggregation in wireless sensor networks can save energy or improve network capacity? In this report, we give a positive answer. Facing on a dedicated application, without lost of generality, we calculate the energy consuming and the capacity in 2 different topologies.

A technical report from ETSI (European Telecommunications Standards Institute) proposed some urban applications to guide further development of smart city based on WSNs [1],e.g. water and gas smart metering, waste management, air pollution monitoring and alerting, acoustic noise monitoring, parking management system etc.

Take water metering for example. The water metering application mostly consists in collecting data from the meters to the wide area network access points that transfer them over cellular or wired networks to the utility company information system. They required every point daily update, and the amount of data is 100 bytes. At the same time, there is a tolerated loss of 10^{-2}.

We consider two network topologies to discuss the energy consumption and network capacity, one is 1-hop network, another is 1D network. In 1-hop network, every node (we set 5 sensors) is directly connected to sink. Assuming the transmission range of a sensor is R, and the distance between every sensor is L and define $L \geq R$ to reduce RF interference, i.e. every two sensors can't communicate. In addition, sensors are set close to the sink to ensure that they are direct communicated and to prevent lost packets during transmission. In 1D network, it's assumed a chain topology. In this model, 5 sensor are placed l away from each other in a straight line and $l \leq R \leq 2l$ to ensure that a chain connection is formed. This two topologies are shown in figure 1.

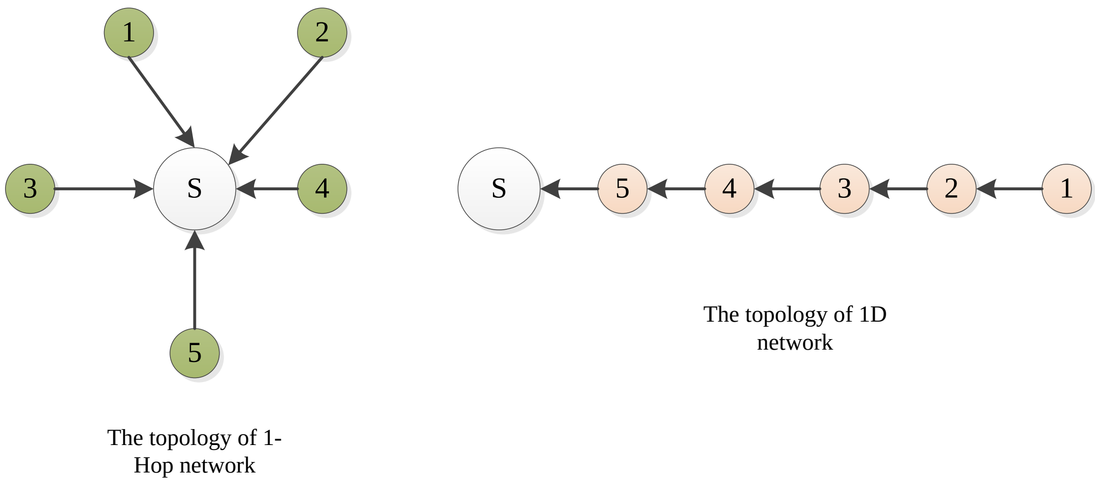


Figure 1: The two different topologies

The circuits radio emitting in the band of ISM 868MHz begin from 19.2kbit/s, and the power losses are considered identical regarding to the power of transmission and receive, i.e. $P_{tx} = 62.5mW$ and $P_{rx} = 53.7mW$.

We consider the sensor 5 in two topologies, because in 1D network, the sensor 5 is easily become the bottleneck for the whole network, and in 1-hop network, every sensor is similar with each other. In others words, we all consider the worst case in two topologies.

We assume a collection of two-day data in water metering application and compare the energy used and the capacity required with and without aggregation. In 1-hop network, sensor 5 needs to report 2 packets during the 2 days, and consuming energy as $2 \cdot P_{tx}$. In contrast, with aggregation, the 2 days data maybe highly temporal correlated and due to the tolerated loss, we just need one packet to forecast or recovery the second packet, thus the energy consuming is P_{tx}. The aggregation procedure can be recovered by sink, the computation in sensor can be negligible. Therefore, after aggregation, in 1-hop network, the saving energy can reach 50%. Similarly, in 1D network, beside transmitting the own data to sink, sensor 5 needs to receive the other 4 sensors' data and forward to sink if there is no aggregation scheme. The process consume more energy as $2 \cdot (4 \cdot P_{rx} + 5 \cdot P_{tx})$. In terms of aggregation, every packet can be aggregated in its next hop, so sensor 5 just needs to transmit one aggregated packet per day, thus the energy consuming is $2 \cdot (P_{tx} + P_{rx})$. In a better situation, second packet can be forecast or recovered by the first packet, the energy consuming is just $P_{tx} + P_{rx}$. The saving energy can reach 79%, as shown in figure 2.

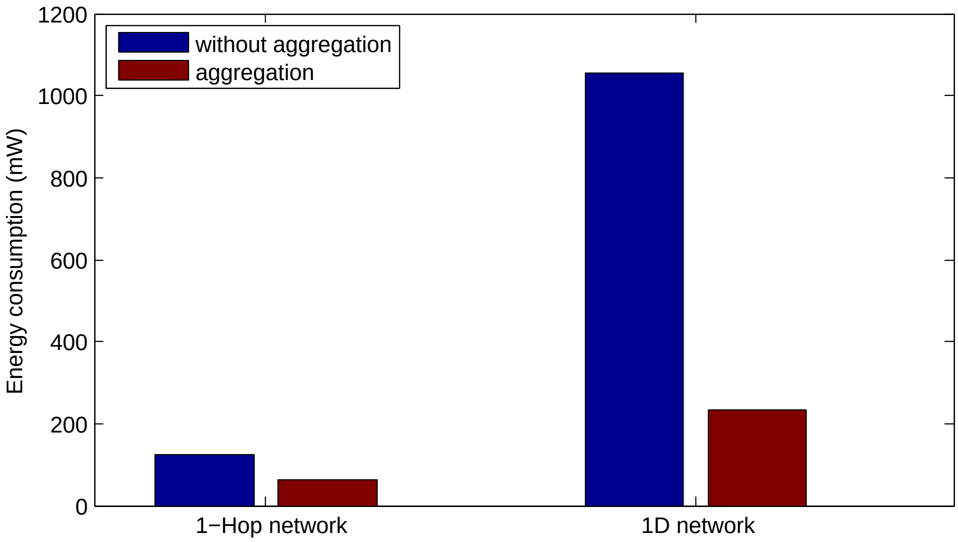


Figure 2: The energy consumption in two different topologies

We use the amount of packet size to estimate the capacity required. As mentioned above, water metering application needs a packet one day, and the amount of packet is 100 bytes, and also for collecting 2-days data. In 1-hop network, sensor 5 needs to report 2 packets, i.e. 200 bytes; in contrast, it only report 1 packet with aggregation, which saves the capacity by 50%. In 1D network, sensor 5 needs to forward 4 packets every day, the capacity from sensor 5 to sink is 500 bytes one day. With aggregation, the packets will be aggregated together, sensor 5 only sends 2 packets one day, and the second-day packet can be forecast or recovered by aggregation scheme, which can save the capacity by 80%, as shown in figure 3.

After analysis of single sensor, we also calculate the total energy consumption and capacity regarding the whole network. In the above paragraph, we assume the best situation with aggregation, here we will discuss the relationship between energy (capacity) and the aggregation ratio[1], and we assume sensors need report during 30 days.

In 1-hop network, every sensor can't be communicated with each other, thus we proposed using the temporal correlation to aggregate the data. Temporal correlation can be discovered in the consecutive time series, we set $w_{1h} \in [0.1, 1]$ as the aggregation ratio in the 1-hop scenario. The total packets is 30, when $w_{1h} = 0.1$, the required packets is only $30 \cdot w_{1h} = 3$, which means

[1] we define the aggregation ratio as $w \in (0, 1]$, and $w = \frac{n}{N}$ where n is the really transmitted packets due to aggregation, while N is the total packets. Thus there only w% of the generated packets are really transmitted. $w = 1$ means there is no correlation between packets. The smaller w is, the smaller number of required packets, the higher correlation, the higher aggregation degree.

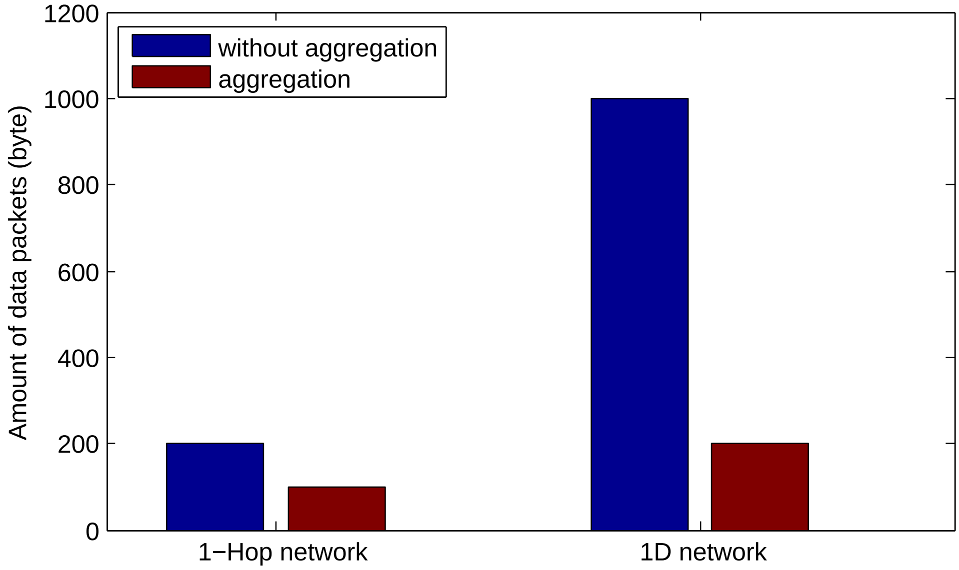
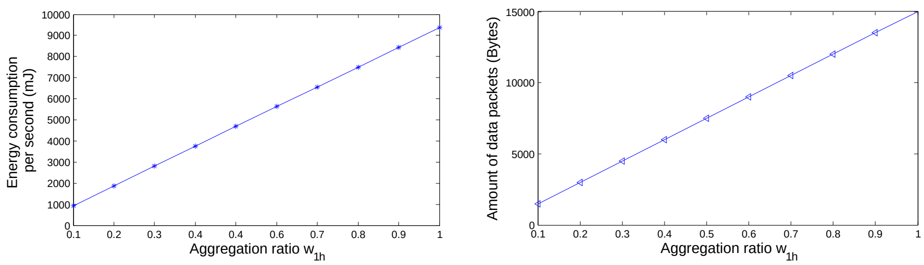


Figure 3: the network capacity in two different topologies

10 packets can be aggregated together. In contrary, $w_{1h} = 1$ means there is no correlation between packets, and the energy consumption can be written as $5 \cdot P_{tx} \cdot 30 \cdot w_{1h}$. Similarly, the capacity consuming is $500 \cdot 30 \cdot w_{1h}$, as show in figure 4 (a) and (b). It is obviously that the higher temporal correlation, the less the aggregation ratio, the whole network consumes less energy and saves more capacity.

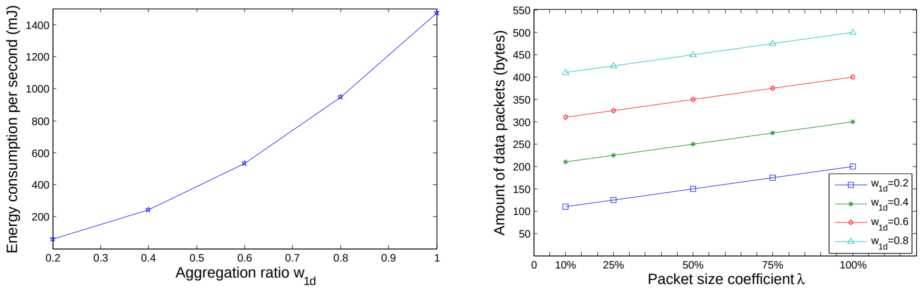
In 1D network, we can adjust the distance between nodes to make them geographically closer for the spatial correlation. If there is no correlation in nodes, the energy consumption for the network is $\sum_{i=1}^{5} i \cdot P_{tx} + (i-1) \cdot P_{rx}$, where i is the sensors' number. We set aggregation ratio $w_{1d} \in [0.2, 1]$, thus the energy consumption can be written as $\sum_{i=1}^{5w_{1d}} i \cdot P_{tx} + (i-1) \cdot P_{rx}$, as shown



(a) The energy consumption with different aggregation ratio w_{1h}

(b) The capacity with different aggregation ratio w_{1h}

Figure 4: Energy and capacity consumption with different aggregation ratio w_{1h} in 1-hop network



(a) The energy consumption with w_{1d} (b) The link capacity with different w_{1d} and λ

Figure 5: Energy and capacity consumption with different aggregation ratio w_{1d} in 1D network

in figure 5 (a).

With regard to the network capacity in 1D network, data aggregation scheme may change the packet size (we set packet size coefficient λ in section 2.4). Using $\lambda = 10\%, 25\%, 50\%, 75\%$ and 100% as the packet size increment. The maximum link capacity for one day without aggregation ($w_{1d} = 1$) is 500. And when $w_{1d} < 1$, the capacity will be formulated as:

$$C_{link_{1d}} = 100 \cdot 5 \cdot w_{1d}(1 + \lambda)$$

We plot the relationship between capacity and aggregation ratio in figure 5(b). Overall, proportional to the aggregation degree, network capacity is saved; that is to say that aggregation scheme indeed improve the network capacity. Where we need to concern is when $w_{1d} = 0.8$, the required packets is 4, i.e. sensor 1 and 2 can be aggregated in one packet, the capacity will be as same as no aggregation scheme with $\lambda = 100\%$; but along with the aggregation ratio decreasing, the capacity is saved by the aggregation scheme. In other words, we need to trade off the packet size change and aggregation ratio when using aggregation (we discuss the trade off in section 2.4).

2.2 Routing layer-benefits from aggregation

In the context of WSNs, routing protocol should save energy and, then, extend the network lifetime, (e.g. Optimized Link State Routing (OLSR)[18], Greedy Perimeter Stateless Routing(GPSR)[21], Gradient-based Routing(GBR)[41], and Simple Random Walk Routing(SRW)[30]). Note that, as data aggregation, the goal of such routing protocol is also to save energy. However, by modelling the energy consumption for the mentioned protocols, we show that data aggregation is always the most basic solution to save energy.

There are different models proposed for modelling energy consumption in battery-powered wireless networks. Here we use the model described in [20] [23] as the model for energy consumption for both transmission and reception, and formulated as:

$$E_b = E_{T_x} + \Gamma \times E_{R_x} \tag{1}$$

Where E_{T_x} is the energy consumed to transmit one bit of information while E_{R_x} is the energy consumed to receive the same bit of information at targeted receivers; and Γ is the neighbours' number which need to report to the receiver. Actually, E_b denote the total energy cost of a single bit in one hop transmission, including transmission and reception cost.

OLSR [18] is one of the proactive routing protocol which is used in wireless mobile ad hoc networks. It consists to periodically exchange topology information in Topology Control (TC) message in order to establish a route to any destination. Multipoint relays is used to optimize TC message flooding. The respective energy cost is as follow. Cost of **Hello(H) message** is expressed as[23]:

$$E_H = f_H \times S_H \times E_b$$

where f_H and S_H are respectively the frequency and the size of the packet of H message(in the following, we use f and S respectively denotes the frequency and size of packet). And similarly, the cost of **TC message** is[23]:

$$E_{TC} = f_{TC} \times S_{TC} \times N_{TC} \times E_b$$

where the N_{TC} is the average number of multipoint. As the transmission of data packet in OLSR (such as GPSR, GBR and SRW) protocol is unicast mode, the cost of **unicast data message** is[23]:

$$E_D = f_D \times (S_D + S_{ACK}) \times L_{OLSR} \times N_{T_x} \times E_b$$

where S_{ACK} is the size of acknowledgement packet and N_{T_x} defined as the average number of retransmission required before the successful transmission in realistic channel condition; and the L_{OLSR} is the average number of hops in OLSR. Thus the total energy cost in OLSR is[23]

$$E_{OLSR} = E_H + E_{TC} + E_D \triangleq f_{OLSR}(E_b)$$

where we defined a function f_{OLSR}, and we find the energy cost in OLSR is a proportion function of E_b.

GPSR [21] is the most will known beacon-based geographic routing. In GPSR, nodes periodically send *hello messages* in order to get its 1-hop nodes location. To send data to the destination, at each hop forwarding node selects its neighbour which will minimize the distance to the destination. the total energy cost is [23]

$$E_{GPSR} = E_H + E_D + E_{PU} \triangleq f_{GPSR}(E_b)$$

where E_{PU} is the energy cost of locations updating, we can consider the value of updating is a *constant*, thus the energy cost is a proportion function of E_b.

In simple Gradient-Based Routing(GBR)[41], only the control packet, *Advertisement message (ADV)* is periodically broadcast in the network. The cost of **ADV message** is [23]

$$E_{ADV} = f_{ADV} \times S_{ADV} \times N_D \times N_s \times E_b$$

where N_D is the number of nodes in the network and N_s is the number of sink in the network. Thus, the total energy cost in GBR is [23]

$$E_{GBR} = E_{ADV} + E_D \triangleq f_{GBR}(E_b)$$

similarly, the energy cost in GBR is a proportion function of E_b.

A simple random walk (SRW)[30] routing is defined as where nodes use *hello message* to set their 1-hop neighbors information. Then, when node has data to send, it randomly selects the next-hop among its 1-hop neighbour before relaying the data to the selected node. The total energy cost is[23]

$$E_{SRW} = E_H + E_D \triangleq f_{SRW}(E_b)$$

the energy cost in SRW is a proportion function of E_b.

Comparing these four routing protocols, although they use different methods to design the routing (e.g. proactive routing, gradient routing or geographic routing), the energy consumption is more or less similar, and all the energy consumption is a proportion function of E_b. Regarding E_b, in equation 1, we know the affect factor is E_{T_x} and E_{R_x}. The purpose of aggregation is exactly reduce the redundant information; the spatial aggregation can be used to reduce the spatial correlated data (in section 3.1, it can primarily reduce the reception cost E_{R_x}), and the temporal aggregation is used to reduce the temporal correlated data (in section 4.2 which can substantially reduce the transmission energy cost E_{T_x}). The energy efficient routing protocols just adjust the indecisive factors(e.g. change broadcasting the flooding message (OLSR) as one control message (GBR) in network etc), however the fundamental factor should be the energy consumption in transmission and reception.

We use aggregation ratio w to describe the extent of aggregation scheme. Thus, assuming we use the aggregation scheme in different routing protocols, the function of energy cost can be redefined as:

$$f'_{OLSR}(E_b) = wf_{OLSR}(E_b) = wE_{OLSR}$$

$$f'_{GPSR}(E_b) = w(E_H + E_D) + E_{PU}$$

$$f'_{GBR}(E_b) = wf_{GBR}(E_b) = wE_{GBR}$$

$$f'_{SRW}(E_b) = wf_{SRW}(E_b) = wE_{SRW}$$

As we mentioned above, aggregation ratio w is a fraction no more than 1, thus the total energy cost would be decreased.

We compare the energy cost in different routing protocols, and in the scenario of SRW, we set aggregation ratio w from 0.2 to 1 show the benefits form aggregation. In figure6, we can see that the energy consumption in the four routing protocols are really different but the difference is not significant as the effect of aggregation scheme; since they just adjust the unicast or broadcast message size or some other indecisive factors. We use aggregation ratio w in the situation of SRW protocol in figure 6, when the $w = 1$, which means all the generated packets are required, i.e. there is no aggregation, the energy consumption is the same as original SRW protocol. With the decrease of w, the required packets is gradually smaller, thus the energy cost is correspondingly decrease. At the meanwhile, we can see from the figure 6 that, routing protocol SRW consume the smallest energy in the 4 protocols, and it can save energy comparing with GPSR; while the saving energy are more and can be achieve 80 % in SRW when $w = 0.2$. It means appropriate aggregation scheme is more useful and efficient to reduce the energy cost than routing protocols. Therefore, data aggregation is the basic and fundamental solution to reduce the energy waste in wireless sensor networks.

2.3 MAC layer-benefits from aggregation

MAC protocols, specially in WSNs, are used to share the medium and avoid collisions between neighbours. Because sensor nodes have low computational, synchronisation capabilities and also memory capacities, MAC protocols face to several limitations[3]. Generally speaking, MAC protocols for WSNs usually use a duty cycle[33] mechanism to save energy. Since the receiving, sending and listening energy costs are approximately the same for usual radio chips (the power consumption of node components is shown in figure 7), the only way to save energy is to turn off the radio (e.g. to switch to sleep mode). The basic idea of duty-cycle MAC protocols consists in a alternatively wake up node and switch to sleep mode; The differences between these protocols are just about the preamble length, the type of configuration in the control packet, and the

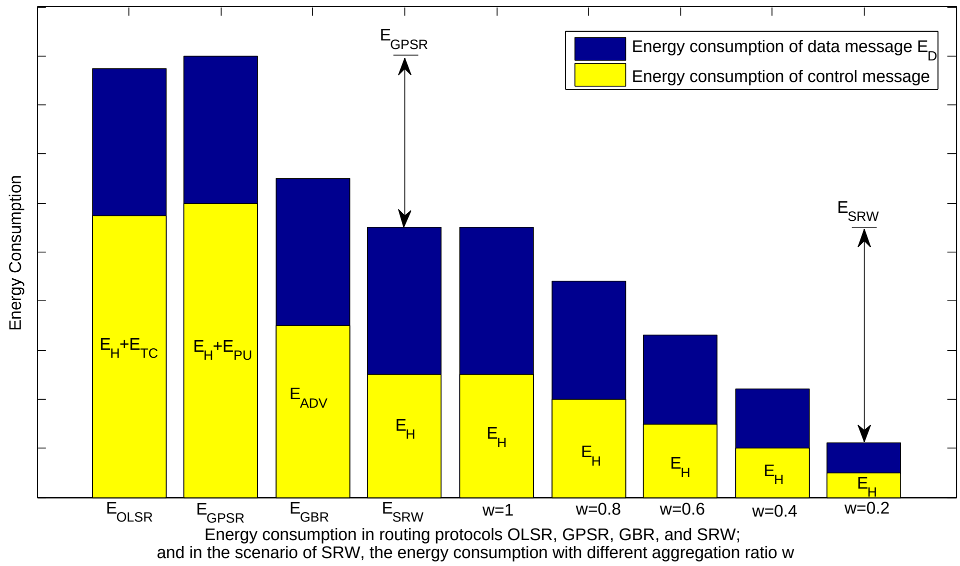


Figure 6: The energy consumption comparison in different routing protocols with and without aggregation

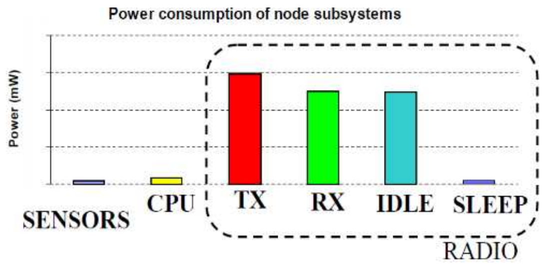


Figure 7: Power consumption of typical node components

convective window. Thus, there are different MAC protocols for different purposes or solving different problems, such as BMAC[33], XMAC[7], SMAC[42] etc.

In BMAC[33], nodes pick a random wakeup time and then alternately sleep and wakeup. BMAC uses LPL (Low Power Listening) method which consists in listening periodically to the channel for radio activity. When a node needs to send a message, it sends a preamble (sequence of bits) which duration is equal to the duty-cycle period. Each node, when it wakes up, senses the channel, if it detects the preamble it stays awake until the end of the communication in order to receive the packet, otherwise it goes back to sleep. The main advantage of this protocol is that there is no need for a global time synchronization. Nevertheless the length of the preamble and the overhearing problem (nodes listening the preamble even if they are not the destination of the packet) can leads to high energy consumption.

Enhancements have been propose by XMAC[7] for instance. They aim at reducing energy consumption by avoiding unnecessary listening. In XMAC, the long preamble is replaced by the strobe preamble. The receiver wakes up randomly, when it listens a preamble and it's the receipt, it will send a ACK frame to sender, and completes the whole process of transmission. Otherwise, it will sleep for the defined cycle.

SMAC[42] makes the nodes in a neighbourhood synchronized in order to wake up at the same time and access the medium by contention using carrier sense and RTS/CTS mechanisms. We show the different MAC mechanisms in figure 8, which describes the whole process of successful sending-receiving in the 3 MAC protocols.

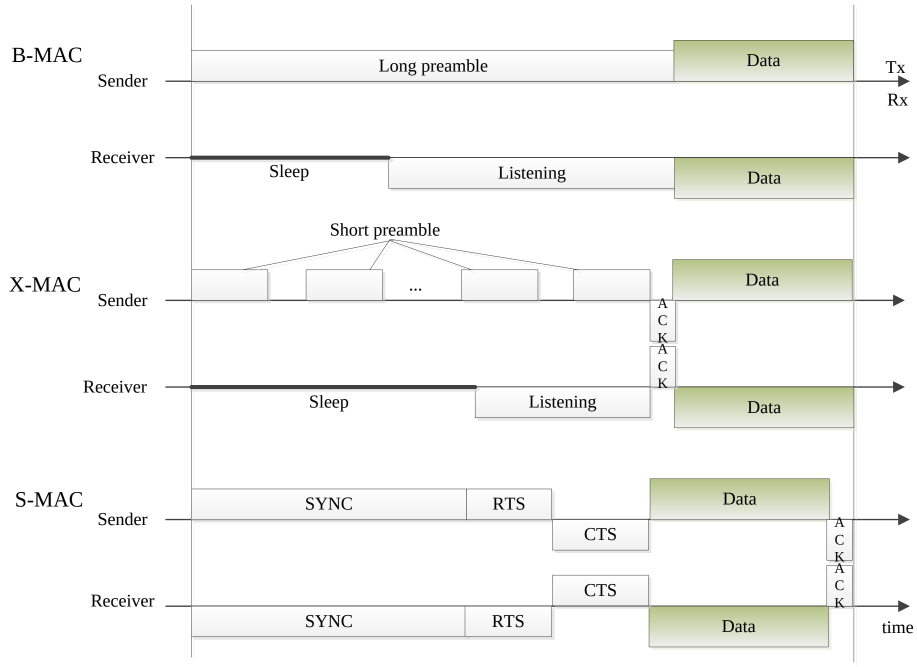


Figure 8: Schematic comparison of the timelines between B-MAC, X-MAC and S-MAC

From figure 8, we can model the energy consumption in the 3 protocols; without loss of generality, we also set Γ as the number of receivers. And to make the model more clearly, we use the parameters in table 1 to define the energy consumption model for different protocols.

For B-MAC, successfully transmit a packet will consume E_{BMAC}, which can be formulated as follow:

$$\begin{aligned} E_{BMAC} &= preamble\, energy + sending\, energy \\ &\quad + \beta \cdot \Gamma \cdot (listen\, energy + sleep\, energy) \\ &= P_{tx} d_p + P_{tx} d_{data} + \beta \cdot \Gamma \cdot (P_s d_s + P_{rx} d_l + P_{rx} d_{data}) \end{aligned}$$

While X-MAC uses strobe preamble to reduce the energy consumption of the long preamble, which has a expected number of iterations required to determine the preamble frequency. Thus,

$$\begin{aligned} E_{XMAC} &= (preamble\, energy + ACK\, listen\, energy) * (expected\, repetition\, required) + sending\, energy \\ &\quad + \beta \cdot \Gamma \cdot [lisen\, energy + sleep\, energy + receiving\, energy + sending\ ACK\ energy] \\ &= (P_{tx} d_p + P_{rx} d_{ACK}) * \alpha + P_{tx} d_{data} + \beta \cdot \Gamma \cdot (P_{rx} d_l + P_s d_s + P_{rx} d_{data} + P_{tx} d_{ACK}) \end{aligned}$$

Before sending packets to the receiver,S-MAC needs to synchronize the neighbours. The energy

Table 1: Model parameters and value

Parameter	Definition	Value		
		BMAC	XMAC	SMAC
P_{tx}	the power required for transmission	62.5	–	–
P_{rx}	the power required for receive	53.7	–	–
P_s	the power required for sleep	0.02	–	–
d_p	the duration for preamble	60	7.8	
d_s	the duration for sleep	22.92	42.8	
d_l	the duration for listening	37.08	10	
d_{data}	the duration for data transmission	15	–	–
d_{ACK}	the duration for acknowledgement		7.2	7.2
d_{SYNC}	the duration for synchronization			25
d_{RTS}	the duration for Request to Send			13.9
d_{CTS}	the duration for Clear to Send			13.9
α	short preamble repetition required in XMAC		4	
Γ	number of neighbours	20	–	–
β	the probability of successful receiving a packet	50%	–	–

"–" show the value is the same with left column.
The units for P_{xx} is mW, and for d_{xx} is ms.

consumption is E_{SMAC}, which can be formulated as:

$$\begin{aligned}E_{SMAC} &= SYNC\,energy + RTS\,energy + sending\,energy + receiving\,CTS + receiving\,ACK\\&\quad + \beta \cdot \Gamma \cdot (SYNC\,energy + CTS\,energy + receiving\,energy + ACK\,energy)\\&= P_{tx}d_{SYNC} + P_{tx}d_{RTS} + P_{tx}d_{data} + P_{rx}d_{CTS} + P_{rx}d_{ACk}\\&\quad + \beta \cdot \Gamma \cdot (P_{rx}d_{SYNC} + P_{tx}d_{CTS} + P_{rx}d_{data} + P_{tx}d_{ACk})\end{aligned}$$

In table 1, we calculate the value for 3 protocols to compare the energy consumption. We set transmission power is $P_{tx} = 62.5mW$, receiving power is $P_{rx} = 53.7mW$, sleeping energy is $P_s = 0.02mW$, and bandwidth is $19.2kbps$. We assume the total time for successful transmission is the same, and data packet is 36 Bytes. Thus the d_{data} in 3 protocols is the same, $15ms$. Based on the parameters given in [7], we can calculate d_p, d_{ACK}. Because we assume the time for transmission is the same in 3 protocols, given the expected repetition required number $\alpha = 4$, we can get the long preamble is $60ms$ (i.e. $d_p = 60ms$ in BMAC). In XMAC, once receiver listen a integral preamble, it will send ACK in the end of preamble. Thus, the duration for listening should meet $d_p \leq d_l < 2 \cdot d_p + d_{ACK}$, so $d_l = 10ms$ and $d_s = 42.8ms$. In BMAC, the receivers are randomly wake up, without loss of generality, we use golden section ratio to set the value, thus $d_l = 37.08ms$ and $d_s = 22.92ms$.

SMAC uses time slots to synchronize and transmit. Due to the parameters in [42], listen interval $75ms$ can be divided to 30 slots, and 10 for SYNC, 20 slots for data. And summing ACK duration is the same with XMAC, we can calculate all the parameters for energy consumption.

Assuming we need 100 packets in a given duration, the neighbours' number is 20, and the probability for successful receiving a packet is 50%. We plot the energy consumption in figure 9. We can see, the energy consumption among different protocols are different, which means appropriate MAC protocol indeed saving energy.

While assuming there is an aggregation scheme which make the aggregation ratio $w = 0.5$, i.e. there are only 50% of the generated packets need to be transmitted. For the 3 protocols, the

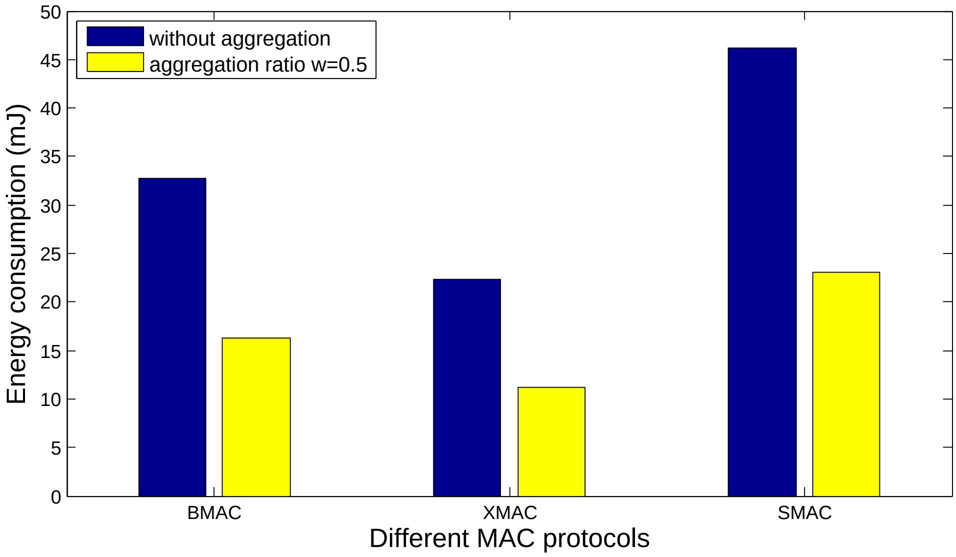


Figure 9: The energy consumption comparison in different MAC protocols with and without aggregation, where aggregation ratio $w = 0.5$

energy consumption decrease 50%, as shown in figure 9. And the benefit for SMAC is significant. Without aggregation, SMAC consumes more than twice than XMAC, while after aggregation, the energy considerably decrease, and the energy saving is more than the difference between SMAC and other protocols. That is to say, if there is a efficient aggregation scheme, there is no need to consider which MAC protocol is better, anyone can achieve a better performance.

Therefore, for the unique application, we can appropriately adjust the MAC parameter due to the aggregation scheme, which will not only decrease the sending and receiving energy, but also decrease the overhead in the MAC protocol(e.g. idle listening), and finally save more energy.

2.4 The trade-off for aggregation

As mentioned above, aggregation in WSNs can save energy regardless of routing layer or MAC layer, and also can improve the network capacity. All of the benefits are derived from the nature of aggregation, data aggregation in WSNs is committed to use the temporal or/and spatial correlation to discover the potential relationship between different packets (the correlated packets in general include the redundant information which is no use for the application), and then essentially reduce the redundant information to save energy and save capacity, i.e. reduce the amount of packets, which is basically reduce the possibility of collision, idle listening etc..

However, we mentioned in section 2.1 that data aggregation may change the packet size, because sometimes we aggregate many information just into one packet. Here, we will analyse the relationship between aggregation ratio w, packet size coefficient λ [2] and energy, network

[2] we define packet size coefficient λ as the rate of the packet size changed, p is the bits for original packet, we set $p^{'}$ is the aggregated packet size, thus $\lambda = \frac{p^{'} - p}{p}$. Note, here λ can be negative value if the aggregated packet size is smaller than the original one. And we set $\lambda \in [-1, 1]$, $\lambda = -1$ means there is no packet need to report after aggregation (i.e. $p^{'} = 0$), and $\lambda = 1$ means the new packet size is twice than the original one (If the packet

capacity.

Assuming a network originally generates N packets, and the average size for the packets is p bits, and in MAC layer successfully transmit a bit will consume E_{bit} energy. Thus, to transmit N packets, the energy consuming is

$$E = N \cdot p \cdot E_{bit}$$

After using aggregation scheme in network, the packets may be increased, we assume the increment is λ. And corresponding the packets increment λ. If the aggregation ratio is w, the real transmitted packets is $w \cdot N$, thus, the aggregated energy consuming is

$$E_{agg} = w \cdot N \cdot p(1 + \lambda) \cdot E_{bit}$$

$E_{agg} \leq E$ illustrate that the aggregation scheme indeed decrease energy consumption, i.e.

$$w \cdot N \cdot p(1 + \lambda) \cdot E_{bit} \leq N \cdot p \cdot E_{bit} \tag{2}$$
$$w \cdot (1 + \lambda) \leq 1 \tag{3}$$

In inequality 3, if the packet size coefficient λ is 0 or very small, the determining factor is aggregation ratio w. Therefore, when we consider to use an aggregation scheme in an application, we need to primarily consider the aggregation ratio. Because aggregation ratio w reflects the efficiency of an aggregation scheme, and at the meanwhile, also implicit the ability of saving energy, which is $1 - w \cdot (1 + \lambda)$.

Similarly, the network capacity (C_{net}) and maximum link capacity (C_{link}) also can be formulated as follow. Without aggregation, the maximum link capacity is:

$$C_{link} = N \cdot p$$

And network capacity is:

$$C_{net} = \sum_{i=1}^{N} i \cdot p$$

With aggregation (aggregation ratio is w), the real transmitted packets is $w \cdot N$, and the packets increment is λ, thus the aggregaed maximum link capacity is:

$$C_{link}^{agg} = w \cdot N \cdot p \cdot (1 + \lambda)$$

And the aggregated network capacity is:

$$C_{net}^{agg} = \sum_{i=1}^{w \cdot N} i \cdot p \cdot (1 + \lambda)$$

The maximum link capacity can be saved by $N \cdot p \cdot [1 - w \cdot (1 + \lambda)]$.

We plot the inequality 3 in figure 10 with $w \in (0, 1]\, and\, \lambda \in [-1, 1]$, and we use colours and areas to describe the potential to save energy and capacity. The warm colours are more approach 1 (in area 5), that is to say the potential to save energy and capacity is lower; while the cool colours are more approach 0 (in area 1,2,3), i.e. the potential to save energy and capacity is higher. From figure 10, we can also see that, if the aggregation ratio $w \leq 0.3$, no matter how change the packet size coefficient λ, the minimum energy and capacity savings can reach

size increases more than 100%, maybe it will lead to the packet loss or traffic congestion, which violate the goal of aggregation. Thus we don't consider the situation of $\lambda > 1$).

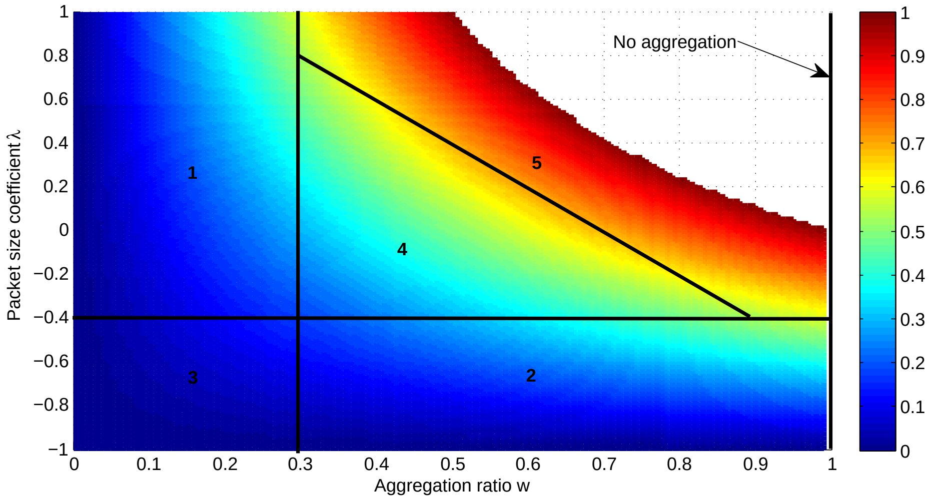


Figure 10: The trade-off between aggregation ratio w and packet size coefficient λ, which show the potential to save energy and capacity.

to 50% (see in area 1 and 3); while if $\lambda \leq -0.4$, the value for saving energy and capacity is also greater or equal 50% with aggregation ratio $w \in (0, 1)$ (see in area 2 and 3). We need to node here that, if $w = 1$ (i.e. there is no aggregation scheme), packet size coefficient λ will be meaningless. And when packet size coefficient $\lambda \leq 0.8$ and aggregation ratio $w \leq 0.9$, i.e. area 4, the minimum energy and capacity savings can get to 30 %. However, in area 5, the ability for saving energy and capacity is not optimistic (just $\leq$ 20%). Thus, we can conclude that, when using aggregation scheme in a real application, we need to trade off the aggregation ratio and packet size coefficient to find the optimal result to save energy and capacity. Therefore, from the perspective of energy and capacity, aggregation scheme is the fundamental solution to save energy and capacity, which make the WSNs live longer and have more capacity to achieve more function. Certainly, we also need to consider the trade off for energy and capacity, if the aggregation scheme is not very efficient(e.g. aggregation ratio $w \propto 1$), the scheme is almost useless for saving energy and capacity. In the following section, we introduce some efficient aggregation scheme: compressing-based and forecasting-based aggregation.

3 Compressing-based aggregation

3.1 Introduction of Compressive Sensing

Shannon sampling theorem [8] defined that the sampling rate should be more than twice the maximum frequency present in the signal. Obviously, this minimum sampling rate is a worst case bound. Recently research found that many natural signals can be transformed to another space, with a small number of the coefficients represent most of the feature of the signals, e.g., audio signals can be transformed into the frequency domain, images can be represented by a discrete fourier transform (DFT) or discrete wavelet transform (DWT).

Compressive sensing (or compressive sample, CS) asserts that certain signals can be recovered from fewer samples than Shannon sampling uses by solving a programming optimization problem form non-adaptive linear projections. The whole process of compressive sensing is shown in fig11.

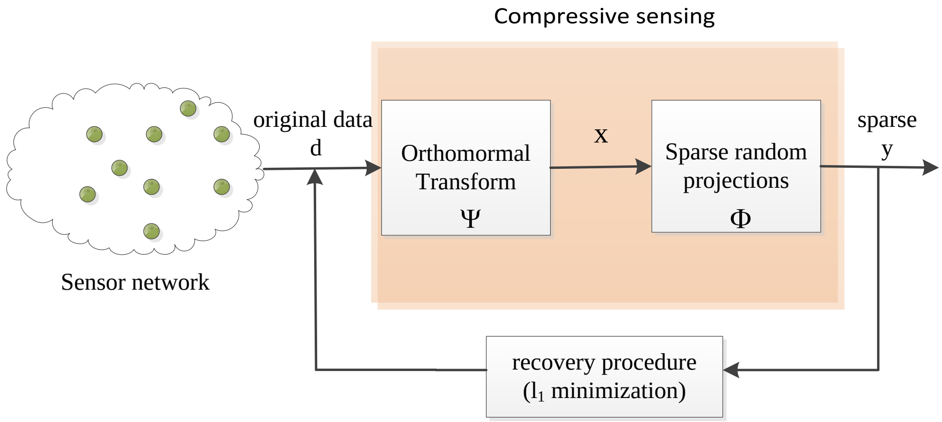


Figure 11: the procedure of compressive sensing

Suppose if a signal $d \in \mathbb{R}^N$can be represented as a sparse signal $x \in \mathbb{R}^N$ in some orthonormal basis $\Psi \in \mathbb{R}^{N\times N}$, the signal can be recovered from M ($M \ll N$)measurements. The sampled signal via CS can be presented as

$$y = \Phi d + e = \Phi\Psi x + e \tag{4}$$

where $\Phi \in \mathbb{R}^{M\times N}$represents a sensing matrix and e is an unknown additive noise during acquisition.

When using CS, it must rely on two conditions : sparsity and incoherence. Sparsity makes it possible to abstract the siagnal with small sample than the Shannon sampling theory used. We say d is k-sparse in the Ψ domain if the number of non-zero coefficient are small and equal to k. And in general, the sensing matrix can be choose randomly, and random matrices are normally incoherent with any fixed basis [8], therefore incoherence between the sensing matrix Φ and transform basis Ψcan be achieved.

There is a usual criterion restricted isometry property (RIP) to detect if the sensing matrix is satisfied to recover the sparse signal, and due to RIP, we know when the number of measurement M satisfies

$$M \geqslant const \cdot k \log \frac{N}{k} \tag{5}$$

Thus, $\mathcal{O}(k \log \frac{N}{k})$ random measurement are enough to recover a signal (when the signal is k-sparse). And in general, we chose $M = 3k \sim 4k$ as the number of measurements.

The recovery procedure is a linear program, l_1 minimization is widely used for CS signal reconstruction, and can be written as ($\beta \geqslant 0$):

$$\min_x \quad \frac{1}{2}\|\Phi\Psi x - y\|_2^2 + \beta\|x\|_1 \tag{6}$$

To solve this problem, there are many algorithm to proposed, e.g. convex optimization algorithms and greedy algorithms[9].

3.2 CS in wireless sensor networks

Because of the characteristics of Compressive sensing, the researchers find that which is adaptive to used in wireless sensor networks. First of all, CS is a compressive method, which can be used to data aggregation in WSN. Second, CS theory shift the energy consumption into the decoder, e.g. the sink node in WSNs, which considerably solve or relieve the energy consume in sensor nodes. As known to all, the energy problem is one of the most importance problem in WSNs. At the mean while, CS can help to reduce the traffic, which also be a part of energy saving.

As mentioned above, CS has many good feature for WSNs, there are many researchers do deeply research on compressive sensing in WSNs. Some papers focus on the data representation to improve the accuracy or energy efficiency [11][39][24][12][13];some papers pay attention on the application[32][10][38][5][28][4][34];some papers concern on the relationship between routing and compression[40][25];and some papers committed to research the performance when using CS in WSNs[43][29][17], and so on.

3.2.1 Data representation in CS

In the processing of CS,there is a measurement matrix $\Phi(M \times N)$to make the data sparsity. And in general, we use random generator to randomly select the projections for WSNs. However, the authors proposed in [11] concern on collect as much information with as little energy as possible. In the beginning, each sensor randomly send its reading to sink (where sink has the whole view of the network), and then when the sink is not satisfied with the data field or the data accuracy;it will determine the projection vector and the path to receive the value. After this, sink will send messages along the path and waits for the projection value; and then update the estimate of the new data field and determine its accuracy. Suppose p is the projection vector,the reduction of differential entropy $\triangle H(p)$ and energy (measured by the numbers of transmissions)$E(p)$ are be used to determine which should be the better projection. To choose the p which make $\frac{\triangle H(p)}{E(p)}$ minimum, they proposed a heuristic algorithm. This method can be used small-scale network since sink need to exactly know the specific field, however the energy consumption may exceed expectation due to the messages dissemination.

In [39], the authors presented distributed algorithm based on sparse random projections to reduce the communication cost of pre-processing the data. To compute one random projection, every sensor j locally generates a random variable Φ_{ij}. If and only if the random variable is non-zero, sensor j sends the value of Φ_{ij} with its own data d_j to the corresponding sensor i. The receiver just simply stores the random projection coefficient$\sum_{j=1}^{n} \Phi_{ij} \cdot d_j$. This process is repeated until every sensor has stored a random projection coefficient. Then sink can query any $\mathcal{O}k \log \frac{N}{k}$ sensors by the idea of compressive sensing to recovery the data. This method reduce the traffic in network since they store the projection in every node, however, every sensor need to consume storage for the purpose.

[24] proposed a simply projection method, which make sensors use projection with the scheduled time slot;it means node $1, 3, 5$ send the reading to sink in slot t_i, and the node $2, 4$ report in t_{i+1}. This method naturally uses the fast fading wireless channels for generating random projections, is simple to implement. However the recovery accuracy may not be guaranteed.

3.2.2 Applications based on CS

In terms of the application using CS, [28]is the first complete design to apply CS theory to data gathering for large-scale wireless sensor networks. They proposed data can be transmitted along with compressed by CS theory, which reduce the possibility of bottleneck (as mentioned in section1 page4, sensor 5 in 1D network is easily become the bottleneck if without aggregation) and present aggregation can improve the network capacity.[32] use Bayesian compressive sensing for sparse event detection, which iteratively estimate the probable values from received data, i.e. using Bayesian compressive sensing to replace l_1 linear minimization. In [38], the authors proposed an intelligent compressive sensing, which introduce autoregressive AR model into the reconstruction of the sensed data, the data accuracy can be improved. The reconstruction is formulated based on equation (6) and can be written as:

$$\min_{x} \quad \alpha \sum_{i}^{N} (d_j - \sum_{j \in S_j} a_{ij} d_{ij}) + \frac{1}{2} \|\Phi \Psi x - y\|_2^2 + \beta \|x\|_1 \tag{7}$$

where the S_i is the one-hop neighbor index set of the i-th node, which is also the support set of the AR model,$a_{ij}$is the j-th AR parameter for the i-th node, and $d_{ij}$is the data of the j-th neighbour for the i-th node. And this equation has been solved in iterate way proposed in [38].

The authors proposed that a compressive sensing process should include random projection, CS construction, and the sampling frequency feedback in [10]. In terms of the sampling frequency feedback, it means setting a parameter which enable the sensor to adjust its sampling rate to keep the reconstruction quality within an acceptable range, called SRI (sampling rate indicator) feedback. It uses L additional reserved samples to evaluate the reconstruction performance. The sink compares the reserved data with the corresponding reconstructed data for calculating the reconstruction quality indicator (RQI), defined as

$$RQI = \frac{\|d_j - y_j\|_2^2}{\|y_j\|_2^2} \tag{8}$$

where j is the index set of the reserved samples,$d_j \in \mathbb{R}^L$ is a part of reconstructed signal corresponding to the reserved samples, and $y_j \in \mathbb{R}^L$ is the vector of reserved samples. If the RQI is below (above) the acceptable reconstruction quality, the sink will send an SRI message to make sensor increase (decrease) the sampling.

[34] and [31] combined the compress sensing and PCA(Principal Component Analysis) to recovery the original data. The PCA [19] is used to find transformations that can sparsify the signal, which originally transform the correlated data to uncorrelated data. In [34]and[31],they use the mean value vector $\bar{x}$ and covariance matrix $\widehat{\Sigma}$ to recovery the original data. In equation (4), d is the original data, and set d_k is a vector of data in time k. Given sample mean vector $\bar{d}$ and the sample covariance matrix $\widehat{\Sigma}$ as:

$$\bar{d} = \frac{1}{K} \sum_{k=1}^{K} d_k$$

$$\widehat{\Sigma} = \frac{1}{K} \sum_{k=1}^{K} (d_k - \bar{d})(d_k - \bar{d})^T$$

According to the PCA theory, the sparse vector s can be written at time k as:

$$s_k = U_N^T (d_k - \bar{d})$$

where U_N is orthonormal matrix and defined as the matrix whose columns are he first M eigenvectors of $\widehat{\Sigma}$, and $U_N U_N^T = I_N$ (I_N is the $N \times N$ identity matrix). Hence, the above equation can be rewritten as:

$$d_k - \bar{d} = U_N s_k = \Psi s_k$$

where the transformation matrix Ψ is set equal to U_N. Using equation (4) (ignoring the white noise e) and above, we can receive:

$$y - \Phi \bar{d} = \Phi(d_k - \bar{d}) = \Phi U_N s_k$$

$$d_k = \bar{d} + U_N s_k$$

From this recovery procedure, the authors use sample mean $\bar{d}$ and covariance matrix $\widehat{\Sigma}$(obtain U_N) to recover the original data. Using this method, firstly need to collect enough data to compute the parameters, and then collect data as CS.

3.2.3 Routing in CS

In terms of the relationship between routing and compression, the authors in [40] present two aggregation mechanisms within CS, one is plain CS aggregation, which means force every link to carry k samples, another is hybrid CS aggregation, i.e., starting CS coding only when the outgoing samples become no less than k. And the result show the hybrid CS aggregation is more energy efficient. In [25], the authors proposed dividing the network into clusters of adjacent nodes and forcing projections to be obtained only from nodes within a cluster. And the authors pointed out that the joint reconstruction has potentially higher chance to reconstruct signal correctly. At the same time, as energy of basis functions are more evenly distributed over overlapped clusters, which leads to better reconstruction performance with joint reconstruction.

3.2.4 Performance with CS

In terms of the compressive sensing performance in WSNs, [17] introduced the mainly feature within CS. In [29], the authors proposed two different ways (plain-CS and hybrid-CS) of applying CS to WSNs at the networking layer. And proved that applying CS naively (plain-CS) may not bring any improvement, and the hybrid-CS can achieve significant improvement in throughput. Meanwhile, the capacity and delay of data gathering with CS is also researched in [43],which based on [28]. The gathering scheme is based on cell, the first phase is that a cell head is designated to collect the data from the neighbour nodes in the same cell, and second step is gathering from the column, and then forward to the sink. Under this scheme, the capacity can be considerable improved, and within TDMA for scheduling, the delay also be bounded.

4 Forecasting-based aggregation

4.1 Introduction of Forecasting Model

Forecasting-based method in WSNs tend to use mathematical model to forecast (due to the high temporal correlation in time series) and reduce the data reporting frequency. In general, two basic model has been used in forecasting in WSNs, one is Auto Regression Moving Average(ARMA); another is Polynomial regression.

4.1.1 ARMA model

The Auto Regression Moving Average (ARMA)[15] model is a widely-used model for time series analysis. It uses the historical data to develop a model for the prediction of future values. Many environmental physical quantities such as temperature and humidity can be modelled with ARMA model. Hence, ARMA model fits well in WSNs monitoring application. It incorporates two terms, the Auto-Regression(AR) term, and the Moving Average (MA) term.

The AR term is a linear regression which represents the self-deterministic part of the time series. If forecasts the current value x_t with p prior values:

$$x_t = \phi_0 + \phi_1 \times x_{t-1} + \cdots + \phi_p \times x_{t-p} \tag{9}$$

An $AR(p)$ model is characterized by the p coefficients.

The MA term captures the influence of random shocks which is independent from autoregressive process. The model consists of random shocks on q prior elements:

$$x_t = \theta \times \epsilon_{t-1} + \cdots + \theta_q \times \epsilon_{t-q} \tag{10}$$

A $MA(q)$ model is characterized by the q coefficients.

The order of an ARMA model is defined as ARMA(p,q); the higher the order is, the higher the algorithmic complexity of the model parameter estimation. Since ARMA model based on the stationary of time series, not all time series can certainly be stationary; thus there is another model based on ARMA, called Auto Regressive Integrated Moving Average(ARIMA)model, which suppose the integrate value between time series should be stationary.$ARIMA(p, i, q)$ model means the model have p AR terms, q MA terms, and the number of difference is i.

In ARMA(p,q), the order of $p$and$q$ is very important for the model performance. In general, we use information criteria to select the value, i.e. Akaike's Information Criteria (AiC)[2] or Bayesian Information Criteria (BiC)[36].

AiC uses the concept of information entropy to offer a relative measure of the information lost for every approximating model, but BiC tend to choose the highest possible dimension. Both AiC and BiC introduce a penalty term for the number of parameters in the model;the penalty term is larger in BiC than AiC,and BiC is consistent, more tolerant than AiC.

4.1.2 Polynomial regression

Polynomial regression is a form of linear regression in which the relationship between the independent variable x and the dependent variable y is modelled as an nth order polynomial. Polynomial regression fits a non-linear relationship between the value of x and the corresponding conditional mean of y, and has been used to describe non-linear phenomena such as sensor data. Although polynomial regression fits a non-linear model to the data, as a statistical estimation problem it is linear, in the sense that the regression function is linear in the unknown parameters that are estimated from the data. For this reason, polynomial regression is considered to be a special case of multiple linear regression, which can be formulated as:

$$p(x) = p_1 x^n + p_2 x^{n-1} + \cdots + p_n x + p_{n+1} \tag{11}$$

Polynomial regression models are usually fit using the method of least squares. The least-squares method minimizes the variance of the unbiased estimators of the coefficients, under the conditions of the Gauss-Markov theorem[16].

4.2 Forecasting-based aggregation

Forecasting is a alternative idea to do aggregation in WSNs, which using less parameter to predict the next value in sink; don't need transmission until the data accuracy is not available.[27][26][6]do some research in this area.

In [26], the authors use ARIMA in wireless sensor networks. They make a sink-driven method, i.e. firstly sink compute the parameters and then send them to corresponding sensor; and when accuracy can't satisfy the threshold,sink will recompute them. Also, this work don't rely on AiC or BiC since the big computation, they defined a metric C

$$C = \alpha \times MAE + (1 - \alpha) \times r_{tran} \tag{12}$$

where MAE is the Mean Absolute Error defined by the threshold, and the r_{tran} is the radio of the number of samples transmitted over the total number of samples. In this equation, $\alpha(0 \leq \alpha \leq 1)$

is used to trade off between MAE and r_{trans}, i.e. between energy and data accuracy. [26] simple the utilisation of ARIMA in WSNs, however, sink need a preparation phase to collect the data and compute the parameters, which is not dynamic in some cases. And the update process is not very continent.

In contrary, [27] present use adaptive-ARMA (A-ARMA) in WSNs directly. A-ARMA based on ARMA, which means use ARMA model to forecast the next time value to achieve aggregation; the difference is that they reduce the computation in every sensor nodes and don't need pre-computation phases. The basic idea of A-ARMA is that each node computes an ARMA model based on a fixed-size window of W consecutive samples. By merely sending the parameters of the ARMA model to the sink node and possibly further to distant servers for rebuilding data, the temporal correlation of these samples within each window is explored. The model parameters will also be used by the distant server for data forecasting, unless it receives new model updates from the sensor nodes. Each node locally verifies the accuracy of the predicted data values with real collected samples. If the accuracy is adequate according to the given threshold, the node assumes that the server can rebuild the data correctly and there is hence no need to report the data. Otherwise, it computes a new model and communicates the parameters to the sink node so as to adjust the forecasting. In order to reduce the complexity in the model estimation process whilst still achieving a high accuracy, a moving window technique is introduced. This means that the verification is required every time the window moves a step ahead. The adaptive nature of the technique is relying on the use of this moving window.

Regarding polynomial regression, [6] use this method in a project *LiveE*![14]. *LiveE*! project deploys a global infrastructure aiming at collecting and distributing environmental information, which include 106 weather stations across 13 countries.[6] use 25 weather stations and data report rate is every $60s$. The 25 weather stations collect environmental data during a time window W, and then the stations determines the coefficients of the polynomial with degree n that fits the collected data in a least squares method. The adaptive n can be sent to server when root mean squared error is lower than a given error threshold. And finally, the stations transmit the corresponding polynomial coefficients to the server, and the original time series can be recovered. Obviously, if the order n is lower than the number of values collected during the time window, the overall data traffic can be significantly reduced. However, this method need the stations to compute the coefficient, if changed as sensor nodes, maybe the computation is too big for them.

5 Compressing v.s. Forecasting

Based on the literature, there are some differences between forecasting aggregation and compressing aggregation. In terms of forecasting-based method, on general, it rely on the time series to forecast, which based on the temporal correlation. While for compressing-based method, it uses the geographic characteristic to compress the data, i.e. using spatial correlation. Forecasting is based on mathematical model to use parameters to predict the next value until the model can't be satisfied the accuracy threshold, i.e. if the network is stable enough or there is no environmental affect, there is no traffic consumption; compressing needs parts of sensor(CS theory) regularly report, and decoder to recovery all the data. Forecasting can be used just 1 sensor network, while compressing is more applicable for large-scale network (the large-scale network is easier to satisfy the requirement of sparsity in CS theory). Regarding the computation, both forecasting and compressing need nodes to simple compute, forecasting need nodes to compute the parameters, while compressing need operation to add the corresponding data together.

For the aggregation ratio, suppose there are N nodes in a network, and the signal is k-sparse, and sink need the network report t times in a given duration, i.e. the generate packets

Table 2: The comparison between A-ARMA and CS

	A-ARMA	Compressive Sensing
Loss?	not relevance	$\surd$
Correlation	temporal	spatial
Computation on sensor	$\surd$	$\surd$
Network scale	≥ 1 sensor	large-scale
Methodology	forecasting aggregation	compressing aggregation
Aggregation ratio	$\frac{1+\eta(t-1)}{t}$	$\frac{3\cdot k}{N}$

number should be $t \cdot N$, and the probability of the unstable sensors is η (suppose every sensor is independent). In this assumption, forecasting aggregated the sensing data into some parameters, which means at beginning, every sensor need to report the parameter, and then just the unstable sensors need to report, i.e. the aggregation ratio is $\frac{N+\eta N(t-1)}{t\cdot N} = \frac{1+\eta(t-1)}{t}$; while compressing need to recovery the original data by the regular reporting, and every time the sensors need $\mathcal{O}(k \log \frac{N}{k})$ measurement out of N (in section3.1, the measurement is usually chose $3 \cdot k$), thus the aggregation ratio is $\frac{p\cdot\mathcal{O}(k \log \frac{N}{k})}{p\cdot N} \approx \frac{3\cdot k}{N}$. It shows that the compressing is not relevant the report times, while if the report frequency is high and the time series is stable relatively, forecasting should be a better choice.

As the representatives of forecasting and compressing, A-ARMA and compressive sensing have the unique feature when realize the aggregation function. Table 2 give the difference and the details of these 2 methods.

6 Conclusion and future work

Data aggregation is a crucial problem in wireless sensor network due to the constrained-energy and constrained-bandwidth. In this report, we proved data aggregation is one of best ways to reduce the energy consuming and improve the network capacity. And then, we present the-state-of-the-art aggregation schemes, the compressing-based and forecasting-based methods. The compressing aggregation focus on compress the data amount accompanied with transmitting based on spatial correlation; while forecasting aggregation tends to use mathematical model to fit the time series and predict the new value due to highly temporal correlation. Different method is committed on different network scale and purpose; it means that compressing is more suitable to the large-scale which need regularly reporting, while forecasting can be used in any network scale which need reporting frequently.

Forecasting and compressing is respectively based on the temporal and spatial correlation, and have their own focus. However, more and more applications need to combine temporal and spatial correlation together to achieve better aggregation efficiency. Thus in the future, we will continue focus on aggregation scheme, at the meanwhile, we will pay attention on whether the traffic model affect the aggregation function, and then tend to propose a unifying aggregation scheme.

References

[1] Spectrum requirements for SRD, M3N and SM applications. *ETSI TC ERM, TR 103 055, v1.1.1*, September 2011.

[2] Hirotogu Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer, 1998.

[3] Abdelmalik Bachir, Mischa Dohler, Thomas Watteyne, and Kin K Leung. Mac essentials for wireless sensor networks. *Communications Surveys & Tutorials, IEEE*, 12(2):222–248, 2010.

[4] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak. Compressive wireless sensing. In *The Fifth International Conference on Information Processing in Sensor Networks, 2006. IPSN 2006.*, pages 134 –142, 0-0 2006.

[5] W. Bajwa, A. Sayeed, and R. Nowak. Matched source-channel communication for field estimation in wireless sensor network. In *Fourth International Symposium on Information Processing in Sensor Networks, 2005. IPSN 2005.*, pages 332 – 339, april 2005.

[6] E. Ben Hamida, H. Ochiai, H. Esaki, P. Borgnat, P. Abry, and E. Fleury. Measurement analysis of the live e! sensor network: Spatial-temporal correlations and data aggregation. In *Applications and the Internet, 2009. SAINT '09. Ninth Annual International Symposium on*, pages 263–266, 2009.

[7] Michael Buettner, Gary V. Yee, Eric Anderson, and Richard Han. X-mac: a short preamble mac protocol for duty-cycled wireless sensor networks. In *Proceedings of the 4th international conference on Embedded networked sensor systems*, SenSys '06, pages 307–320, New York, NY, USA, 2006. ACM.

[8] E.J. Candes and M.B. Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21 –30, march 2008.

[9] Emmanuel J Candes and Justin K Romberg. Signal recovery from random projections. In *Electronic Imaging 2005*, pages 76–86. International Society for Optics and Photonics, 2005.

[10] W. Chen and I.J. Wassell. Energy-efficient signal acquisition in wireless sensor networks: a compressive sensing framework. *Wireless Sensor Systems, IET*, 2(1):1–8, 2012.

[11] Chun Tung Chou, R. Rana, and Wen Hu. Energy efficient information collection in wireless sensor networks using adaptive compressive sensing. In *Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on*, pages 443–450, 2009.

[12] A. Ciancio, S. Pattem, A. Ortega, and B. Krishnamachari. Energy-efficient data representation and routing for wireless sensor networks based on a distributed wavelet compression algorithm. In *The Fifth International Conference on Information Processing in Sensor Networks, 2006. IPSN 2006.*, pages 309–316, 2006.

[13] C.G.N. de Carvalho, D.G. Gomes, J.N. de Souza, and N. Agoulmine. Multiple linear regression to improve prediction accuracy in wsn data reduction. In *2011 7th Latin American Network Operations and Management Symposium (LANOMS),*, pages 1 –8, oct. 2011.

[14] E.B. Hamida, P. Borgnat, H. Esaki, P. Abry, E. Fleury, et al. Live e! sensor network: Correlations in time and space. In *XXIIe Colloque GRETSI-Traitement du Signal et des Images*, 2009.

[15] James Douglas Hamilton. *Time series analysis*, volume 2. Cambridge Univ Press, 1994.

[16] David Harville. Extension of the gauss-markov theorem to include the estimation of random effects. *The Annals of Statistics*, 4(2):384–395, 1976.

[17] J. Haupt, W.U. Bajwa, M. Rabbat, and R. Nowak. Compressed sensing for networked data. *Signal Processing Magazine, IEEE*, 25(2):92 –101, march 2008.

[18] P. Jacquet, P. MÃ¼hlethaler, T. Clausen, A. Laouiti, A. Qayyum, and L. Viennot. Optimized link state routing protocol for ad hoc networks. pages 62–68, 2001.

[19] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.

[20] H. Karl and A. Willig. *Protocols and Architectures for Wireless Sensor Networks*. Wiley, 2007.

[21] Brad Karp and Hsiang-Tsung Kung. Gpsr: Greedy perimeter stateless routing for wireless networks. In *Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 243–254. ACM, 2000.

[22] Kavi K Khedo, Rajiv Perseedoss, Avinash Mungur, et al. A wireless sensor network air pollution monitoring system. *arXiv preprint arXiv:1005.1737*, 2010.

[23] R. D. Komguem, I. Amadou, G. Chelius, and F. Valois. Routing protocols: When to use it in terms of energy? In *IEEE WCNC*, Paris, France, April 2012.

[24] M. Laifenfeld and I. Bilik. Distributed compressive sensing and communications in wireless sensor networks. In *2012 IEEE 27th Convention of Electrical Electronics Engineers in Israel (IEEEI),*, pages 1–5, 2012.

[25] Sungwon Lee, Sundeep Pattem, Maheswaran Sathiamoorthy, Bhaskar Krishnamachari, and Antonio Ortega. Spatially-localized compressed sensing and routing in multi-hop sensor networks. In *Geosensor Networks*, pages 11–20. Springer, 2009.

[26] Chong Liu, Kui Wu, and Min Tsao. Energy efficient information collection with the arima model in wireless sensor networks. In *Global Telecommunications Conference, 2005. GLOBECOM'05. IEEE*, volume 5, pages 5–pp. IEEE, 2005.

[27] Jialiang Lu, Fabrice Valois, Mischa Dohler, and Min-You Wu. Optimized data aggregation in wsns using adaptive arma. In *Sensor Technologies and Applications (SENSORCOMM), 2010 Fourth International Conference on*, pages 115–120. IEEE, 2010.

[28] C. Luo, F. Wu, J. Sun, and C.W. Chen. Compressive data gathering for large-scale wireless sensor networks. In *Proceedings of the 15th annual international conference on Mobile computing and networking*, pages 145–156. ACM, 2009.

[29] J. Luo, L. Xiang, and C. Rosenberg. Does compressed sensing improve the throughput of wireless sensor networks? In *2010 IEEE International Conference on Communications (ICC),*, pages 1 –6, may 2010.

[30] Issam Mabrouki, Xavier Lagrange, and Gwillerm Froc. Random walk based routing protocol for wireless sensor networks. In *Proceedings of the 2nd international conference on Performance evaluation methodologies and tools*, page 71. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007.

[31] Riccardo Masiero, Giorgio Quer, Daniele Munaretto, Michele Rossi, Joerg Widmer, and Michele Zorzi. Data acquisition through joint compressive sensing and principal component analysis. In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pages 1–6. IEEE, 2009.

[32] Jia Meng, Husheng Li, and Zhu Han. Sparse event detection in wireless sensor networks using compressive sensing. In *43rd Annual Conference on Information Sciences and Systems, 2009. CISS 2009.*, pages 181–185, March.

[33] Joseph Polastre, Jason Hill, and David Culler. Versatile low power media access for wireless sensor networks. In *Proceedings of the 2nd international conference on Embedded networked sensor systems*, SenSys '04, pages 95–107, New York, NY, USA, 2004. ACM.

[34] G. Quer, D. Zordan, R. Masiero, M. Zorzi, and M. Rossi. Wsn-control: Signal reconstruction through compressive sensing in wireless sensor networks. In *Local Computer Networks (LCN), 2010 IEEE 35th Conference on*, pages 921–928, 2010.

[35] Maneesha V Ramesh. Real-time wireless sensor network for landslide detection. In *Sensor Technologies and Applications, 2009. SENSORCOMM'09. Third International Conference on*, pages 405–409. IEEE, 2009.

[36] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[37] Rui Tan, Guoliang Xing, Jinzhu Chen, Wen-Zhan Song, and Renjie Huang. Quality-driven volcanic earthquake detection using wireless sensor networks. In *Real-Time Systems Symposium (RTSS), 2010 IEEE 31st*, pages 271–280. IEEE, 2010.

[38] Jin Wang, Shaojie Tang, Baocai Yin, and Xiang-Yang Li. Data gathering in wireless sensor networks through intelligent compressive sensing. In *INFOCOM, 2012 Proceedings IEEE*, pages 603–611, March.

[39] Wei Wang, Minos Garofalakis, and K. Ramchandran. Distributed sparse random projections for refinable approximation. In *6th International Symposium on Information Processing in Sensor Networks, 2007. IPSN 2007.*, pages 331–339, 2007.

[40] Liu Xiang, Jun Luo, and Athanasios Vasilakos. Compressed data aggregation for energy efficient wireless sensor networks, 2011.

[41] Fan Ye, Gary Zhong, Songwu Lu, and Lixia Zhang. Gradient broadcast: A robust data delivery protocol for large scale sensor networks. *Wireless Networks*, 11(3):285–298, 2005.

[42] Wei Ye, J. Heidemann, and D. Estrin. Medium access control with coordinated adaptive sleeping for wireless sensor networks. *Networking, IEEE/ACM Transactions on*, 12(3):493–506, 2004.

[43] Haifeng Zheng, Shilin Xiao, Xinbing Wang, and Xiaohua Tian. On the capacity and delay of data gathering with compressive sensing in wireless sensor networks. In *2011 IEEE Global Telecommunications Conference (GLOBECOM 2011),*, pages 1–5, Dec.

Contents

1 Introduction **3**

2 Aggregation benefits **3**
2.1 Some basic results 3
2.2 Routing layer-benefits from aggregation 7
2.3 MAC layer-benefits from aggregation 9
2.4 The trade-off for aggregation 13

3 Compressing-based aggregation **15**
3.1 Introduction of Compressive Sensing 15
3.2 CS in wireless sensor networks 16
3.2.1 Data representation in CS 17
3.2.2 Applications based on CS 17
3.2.3 Routing in CS 19
3.2.4 Performance with CS 19

4 Forecasting-based aggregation **19**
4.1 Introduction of Forecasting Model 19
4.1.1 ARMA model 19
4.1.2 Polynomial regression 20
4.2 Forecasting-based aggregation 20

5 Compressing v.s. Forecasting **21**

6 Conclusion and future work **22**

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399